Amendments to the Specification:

■ Replace the paragraph beginning at page 9, line 12, with the following amended paragraph:

The term "base contacting positions" refers to the four amino acid positions of zinc finger domains that structurally correspond to amino acids arginine 73, aspartic acid 75, glutamic acid 76, and arginine 79 of SEQ ID NO:21. These positions are also referred to as positions –1, 2, 3, and 6. To identify positions in a query sequence that correspond to the base contacting positions, the query sequence is aligned to the zinc finger domain of interest such that the cysteine and histidine residues of the query sequence are aligned with those of finger 3 of Zif268. The ClustalW WWW Service at the European Bioinformatics Institute (~~http://www2.ebi.ac.uk/clustalw;~~ Thompson *et al.* (1994) *Nucleic Acids Res.* 22:4673-4680) provides one convenient method of aligning sequences.

■ Replace the paragraph beginning at page 15, line 10, with the following amended paragraph:

Homeodomains. Homeodomains are simple eukaryotic domains that consist of a N-terminal arm that contacts the DNA minor groove, followed by three α-helices that contact the major groove (for a review, see, e.g., Laughon, (1991) *Biochemistry* 30:11357-67). The third α-helix is positioned in the major groove and contains critical DNA-contacting side chains. Homeodomains have a characteristic highly-conserved motif present at the turn leading into the third α-helix. The motif includes an invariant tryptophan that packs into the hydrophobic core of the domain. This motif is represented in the Prosite database (see the EXPASY online resource ~~http://www.expasy.ch/~~) as PDOC00027 ([L/I/V/M/F/Y/G]-[A/S/L/V/R]-X(2)-[L/I/V/M/S/T/A/C/N]-X-[L/I/V/M]-X(4)-[L/I/V]- [R/K/N/Q/E/S/T/A/I/Y]-[L/I/V/F/S/T/N/K/H]-W-[F/Y/V/C]-X-[N/D/Q/T/A/H]-X(5)- [R/K/N/A/I/M/W]; SEQ ID NO:77). Homeodomains are commonly found in transcription factors that determine cell identity and provide positional

information during organismal development. Such classical homeodomains can be found in the genome in clusters such that the order of the homeodomains in the cluster approximately corresponds to their expression pattern along a body axis. Homeodomains can be identified by alignment with a homeodomain, e.g., Hox-1, or by alignment with a homeodomain profile or a homeodomain hidden Markov Model (HMM; see below), e.g., PF00046 of the Pfam database or "HOX" of the SMART database (online at EMBL, Heidelberg DE ~~http://smart.embl-heidelberg.de/~~), or by the Prosite motif PDOC00027 as mentioned above.

■ Replace the paragraph beginning at page 15, line 28, with the following amended paragraph:

Helix-turn-helix proteins. This DNA binding motif is common among many prokaryotic transcription factors. There are many subfamilies, e.g., the LacI family, the AraC family, to name but a few. The two helices in the name refer to a first α–helix that packs against and positions a second α-helix in the major groove of DNA. These domains can be identified by alignment with a HMM, e.g., HTH_ARAC, HTH_ARSR, HTH_ASNC, HTH_CRP, HTH_DEOR, HTH_DTXR, HTH_GNTR, HTH_ICLR, HTH_LACI, HTH_LUXR, HTH_MARR, HTH_MERR, and HTH_XRE profiles available in the SMART database (online at EMBL, Heidelberg DE ~~http://smart.embl-heidelberg.de/~~).

■ Replace the paragraph beginning at page 16, line 5, with the following amended paragraph:

Helix-loop-helix proteins. This DNA binding domain is commonly found among homo- and hetero-dimeric transcription factors, e.g., MyoD, fos, jun, E11, and myogenin. The domain consists of a dimer, each monomer contributing two α–helices and intervening loop. The domain can be identified by alignment with a HMM, e.g., the "HLH" profile available in the SMART database (online at EMBL, Heidelberg DE ~~http://smart.embl-heidelberg.de/~~). Although

helix-loop-helix proteins are typically dimeric, monomeric versions can be constructed by engineering a polypeptide linker between the two subunits such that a single open reading frame encodes both the two subunits and the linker.

■ Replace the paragraph beginning at page 16, line 16, with the following amended paragraph:

Computational Methods. The amino acid sequence of a DNA binding domain isolated by a method described herein can be compared to a database of known sequences, e.g., an annotated database of protein sequences or an annotated database which includes entries for nucleic acid binding domains. In another implementation, databases of uncharacterized sequences, e.g., unannotated genomic, EST or full-length cDNA sequence; of characterized sequences, e.g., SwissProt or PDB; and of domains, e.g., Pfam, ProDom (online at Institut National de la Recherche Agronomique, Toulouse, FR, ~~http://www.tooulouse.inra.fr/~~), and SMART (Simple Modular Architecture Research Tool, online at EMBL, Heidelberg DE ~~http://smart.embl-heidelberg.de/~~) can provide a source of nucleic acid binding domain sequences. Nucleic acid sequence databases can be translated in all six reading frames for the purpose of comparison to a query amino acid sequence. Nucleic acid sequences that are flagged as encoding candidate nucleic acid binding domains can be amplified from an appropriate nucleic acid source, e.g., genomic DNA or cellular RNA. Such nucleic acid sequences can be cloned into an expression vector. The procedures for computer-based domain identification can be interfaced with an oligonucleotide synthesizer and robotic systems to produce nucleic acids encoding the domains in a high-throughput platform. Cloned nucleic acids encoding the candidate domains can also be stored in a host expression vector and shuttled easily into an expression vector, e.g., into a translational fusion vector with Zif268 fingers 1 and 2, either by restriction enzyme mediated subcloning or by site-specific, recombinase mediated subcloning (see U.S. Patent No. 5,888,732). The high-throughput platform can be used to generate multiple

microtitre plates containing nucleic acids encoding different candidate nucleic acid binding domains.

■ Replace the paragraph beginning at page 17, line 6, with the following amended paragraph:

Detailed methods for the identification of domains from a starting sequence or a profile are well known in the art. See, for example, Prosite (Hofmann *et al.,* (1999) *Nucleic Acids Res.* 27:215-219), FASTA, BLAST (Altschul *et al.,* (1990) *J. Mol. Biol.* 215:403-10.), etc. A simple string search can be done to find amino acid sequences with identity to a query sequence or a query profile, e.g., using ~~Perl (http://bio.perl.org/)~~ PERL to scan text files. Sequences so identified can be about 30%, 40%, 50%, 60%, 70%, 80%, 90%, or greater identical to an initial input sequence.

■ Replace the paragraph beginning at page 17, line 13, with the following amended paragraph:

Domains similar to a query domain can be identified from a public database, e.g., using the XBLAST programs (version 2.0) of Altschul *et al.,* (1990) *J. Mol. Biol.* 215:403-10. For example, BLAST protein searches can be performed with the XBLAST parameters as follows: score = 50, wordlength = 3. Gaps can be introduced into the query or searched sequence as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17):3389-3402. Default parameters for XBLAST and Gapped BLAST programs are available online at the National Center for Biotechnology Information, National Institutes of Health, Bethesda MD ~~at http://www.ncbi.nlm.nih.gov~~.

■ Replace the paragraph beginning at page 17, line 28, with the following amended paragraph:

Hidden Markov Models (HMM's) representing a DNA binding domain of interest can be generated or obtained from a database of such models, e.g., the Pfam database, release 2.1. A database can be searched, e.g., using the default parameters, with the HMM in order to find additional domains (see, e.g., the ~~http://www.sanger.ac.uk/~~Software/Pfam/HMM_search folder at the Sanger Center, UK for default parameters). Alternatively, the user can optimize the parameters. A threshold score can be selected to filter the database of sequences such that sequences that score above the threshold are displayed as candidate domains. A description of the Pfam database can be found in Sonhammer *et al.,* (1997) *Proteins* 28(3):405-420, and a detailed description of HMMs can be found, for example, in Gribskov *et al.,* (1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.,* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.,* (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.,* (1993) *Protein Sci.* 2:305-314.

■ Replace the paragraph beginning at page 18, line 9, with the following amended paragraph:

The SMART database of HMM's (Simple Modular Architecture Research Tool, online at EMBL, Heidelberg DE ~~http://smart.embl-heidelberg.de/~~; Schultz *et al.,* (1998) *Proc. Natl. Acad. Sci. USA* 95:5857 and Schultz *et al.,* (2000) *Nucl. Acids Res* 28:231) provides a catalog of zinc finger domains (ZnF_C2H2; ZnF_C2C2; ZnF_C2HC; ZnF_C3H1; ZnF_C4; ZnF_CHCC; ZnF_GATA; and ZnF_NFX) identified by profiling with the hidden Markov models of the HMMer2 search program (Durbin *et al.,* (1998) *Biological sequence analysis: probabilistic models of proteins and nucleic acids.* Cambridge University Press~~.; http://hmmer.wustl.edu/~~).